Chapter 1
Recent Trends in Deepfake Detection

Kerenalli Sudarshana
https://orcid.org/0000-0001-9581-6835
GITAM School of Technology, Bengaluru, India

Mylarareddy C.
GITAM School of Technology, Bengaluru, India

ABSTRACT

Almost 59% of the world's population is on the internet, and in 2020, globally, there were more than 3.81 billion individual social network users. Eighty-six percent of the internet users were fooled to spread fake news. The advanced artificial intelligence (AI) algorithms can generate fake digital content that appears to be realistic. The generated content can deceive the users into believing it is real. These fabricated contents are termed deepfakes. The common category of deepfakes is video deepfakes. The deep learning techniques, such as auto-encoders and generative adversarial network (GAN), generate near realistic digital content. The content generated poses a serious threat to the multiple dimensions of human life and civil society. This chapter provides a comprehensive discussion on deepfake generation, detection techniques, deepfake generation tools, datasets, applications, and research trends.

INTRODUCTION

As of October 2020, around 4.66 Billion folks on the Internet accounts for Fifty-nine percent of the world population. Ninety-one percent of the total users accessed the Internet through smartphones (J. Clement, 2020). A Social Network platform is a computer-enabled virtual social environment that constitutes a network of people (Dollarhide, 2020). About 3.81 Billion individuals are on any one of the social networks (Dean, 2020). These platforms enable the members to generate information and share opinions, ideas, tags, and other types of social activities online (Kietzmann et al., 2011). The cyber flocks fooled almost Eighty-six percent of Internet users to spread fake news through the social media platform or publishing media platforms. (Center for International Governance Innovation (CIGI), 2019).

DOI: 10.4018/978-1-7998-7728-8.ch001

Deep learning algorithms solve various complex problems ranging from self-driving cars, online-games, big data analytics, natural language processing, computer vision, and computer-human interaction, to quote a few. One such area is deepfake content generation. Deepfakes are digital content generated by swapping the target person's information with the original to deceive the audience (Westerlund M., 2019). A sophisticated deep learning algorithm, commonly used for dimensional reduction in the computer vision domain, is used for deepfake generation. The auto-encoders (Badrinarayanan V. et al., 2017), GANs (Yang, W. et al., 2019) are commonly used to generate more realistic digital content to distinguish by the human sensory organs. Deepfakes not only capable of content swapping but also can generate novel content (Avatarify, 2020). Some software can create real-time deepfakes, and some require just a still image or just a few seconds of an audio bit to generate the deepfake.

The first deepfake was reported in 2017, where a Hollywood actress's face was swapped with a porn actress. The most famous deepfake video, which went viral, was released in Barack Obama's 2018 video (Bloomberg, 2018). Less powered hardware requirements, low learning curves, technology access to the public are a few reasons for the voluminous deepfake traffic on the Internet. Even the source of the generated content, sometimes, is going to be anonymous. After the 2016 US presidential elections, the detection of such manipulated content attracted academics. The data obtained from the https://app.dimensions.ai (dimensions, 2021) discloses the available information on deepfakes till to date. It is as given in table 1. There are totally 62 policy documents and one clinical trail on deepfakes.

*Table 1. Research trend on deepfakes (Source:*app.dimension.ai*)*

Sl. No	Year of Publication	No. of Published Papers	No. of Datasets	No of Patents Filed	No. of grants
1	Before 2018	04	--	118	18
2	2018	65	1		
3	2019	368	2		
4	2020	1299	5		
5	2021	447[1]	8		

The objectives of the proposed chapter are to:

- Provide a concise overview of deepfake technologies.
- Describe various methods employed for generating the deepfakes.
- Describe the key methodologies used to identify deepfakes.
- Enlist the deepfake generation software tools along with their salient features.
- Explore the datasets that were used to assess the deepfake detection techniques.
- The applications and research trends in deepfake technology.

BACKGROUND

Deepfakes are artificially generated digital media using deep learning methods, wherein the features of an individual in a source content are substituted by someone else (Dirik, 2020). The standard category of deepfakes is video deepfakes. They are often circulated on social media and the Internet. The trend